

## ARTICLE

Received 19 Nov 2013 | Accepted 6 May 2014 | Published 19 Jun 2014

DOI: 10.1038/ncomms5037

OPEN

# Vanillin formation from ferulic acid in *Vanilla planifolia* is catalysed by a single enzyme

Nethaji J. Gallage<sup>1,2,3</sup>, Esben H. Hansen<sup>4</sup>, Rubini Kannangara<sup>1,2,3</sup>, Carl Erik Olsen<sup>1,2</sup>, Mohammed Saddik Motawia<sup>1,2,3</sup>, Kirsten Jørgensen<sup>1,2,3</sup>, Inger Holme<sup>5</sup>, Kim Hebelstrup<sup>5</sup>, Michel Grisoni<sup>6</sup> & Birger Lindberg Møller<sup>1,2,3,7</sup>

Vanillin is a popular and valuable flavour compound. It is the key constituent of the natural vanilla flavour obtained from cured vanilla pods. Here we show that a single hydratase/lyase type enzyme designated vanillin synthase (VpVAN) catalyses direct conversion of ferulic acid and its glucoside into vanillin and its glucoside, respectively. The enzyme shows high sequence similarity to cysteine proteinases and is specific to the substitution pattern at the aromatic ring and does not metabolize caffeic acid and *p*-coumaric acid as demonstrated by coupled transcription/translation assays. VpVAN localizes to the inner part of the vanilla pod and high transcript levels are found in single cells located a few cell layers from the inner epidermis. Transient expression of VpVAN in tobacco and stable expression in barley in combination with the action of endogenous alcohol dehydrogenases and UDP-glucosyl-transferases result in vanillyl alcohol glucoside formation from endogenous ferulic acid. A gene encoding an enzyme showing 71% sequence identity to VpVAN was identified in another vanillin-producing plant species *Glechoma hederacea* and was also shown to be a vanillin synthase as demonstrated by transient expression in tobacco.

<sup>1</sup> Plant Biochemistry Laboratory, Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Thorvaldsensvej 40, Frederiksberg C, DK-1871 Copenhagen, Denmark. <sup>2</sup> VILLUM Research Center 'Plant Plasticity', Thorvaldsensvej 40, Frederiksberg C, DK-1871 Copenhagen, Denmark. <sup>3</sup> Center for Synthetic Biology: 'bioSYNergy', Thorvaldsensvej 40, Frederiksberg C, DK-1871 Copenhagen, Denmark. <sup>4</sup> Evolva A/S, Lersø Parkallé 42-44, 5th floor, DK-2100 Copenhagen, Denmark. <sup>5</sup> AU Flakkebjerg, Danish Centre for Food and Agriculture, University of Aarhus, Forsøgsvej, DK-4200 Slagelse, Denmark. <sup>6</sup> Centre de Coopération Internationale en Recherche Agronomique pour le Développement, UMR PVBMT, 97410 Saint Pierre, La Réunion, France. <sup>7</sup> Carlsberg Laboratory, Gamle Carlsberg Vej 10, Valby DK-2500, Copenhagen, Denmark. Correspondence and requests for materials should be addressed to B.L.M. (email: blm@plen.ku.dk).

Vanilla is the world's most popular flavour principle and used in numerous products. The pods of the climbing orchids, *Vanilla planifolia* and *V. tahitensis* are the source of natural vanilla<sup>1</sup>, although trace amounts of vanillin can be found in a variety of different plant species scattered in the plant kingdom<sup>2</sup>. Vanillin (3-methoxy-4-hydroxybenzaldehyde) is the main flavour component of vanilla extract from cured vanilla pods<sup>1,3</sup>. In high concentrations vanillin is toxic to living cells. In the pod it is produced and stored as non-toxic vanillin glucoside, which upon tissue damage is hydrolysed to form the active defense compound, vanillin. Production of vanillin from the orchids is laborious, slow and costly. Five hundred kilograms of vanilla pods yields only 1 kg of vanillin. Less than 1% of the global vanillin production originates from the vanilla orchids. Instead, the vast majority is produced chemically from fossil fuels or by acid hydrolysis of lignin<sup>4</sup>. A biotechnological solution to vanillin production via heterologous expression of the native vanilla orchid pathway genes in microorganisms has not been possible because the pathway has remained unknown. Vanillin has been produced by microbial bioconversion of substrates structurally related to vanillin<sup>5</sup> as well as from glucose<sup>6</sup>.

Previous studies have shown the conversion of a variety of compounds into vanillin glucoside after administration to *V. planifolia* pods. These studies suggest that vanillin glucoside is derived from phenylalanine, the shikimate pathway intermediates or monomeric lignin precursors<sup>7–12</sup>. Vanillin glucoside and *p*-hydroxybenzaldehyde glucoside, the two most abundant aroma compounds in mature vanilla pods, are structurally similar, and a biosynthetic relationship between the formation of these two compounds has been proposed<sup>12,13</sup>. The necessary chain shortening process of a putative phenylpropanoid precursor was suggested to proceed by  $\beta$ -oxidation or by a coenzyme A (CoA)-dependent non- $\beta$ -oxidative pathway<sup>11,14–17</sup>. *p*-Hydroxybenzaldehyde was reported to be formed by chain shortening of *p*-coumaric acid catalysed by *p*-hydroxybenzaldehyde synthase (4-HBS)<sup>18,19</sup>, which was proposed as a precursor for vanillin glucoside biosynthesis. *p*-Hydroxybenzaldehyde would then need to be hydroxylated at its C3 carbon by a monooxygenase (C3H), *O*-methylated at the 3-OH position by a *O*-methyltransferase (OMT), and finally glucosylated at the 4-OH position by a UDP-glucosyltransferase (UGT).

Other studies have suggested that vanillin is formed from L-phenylalanine via the monomeric lignin precursors: cinnamic acid, *p*-coumaric acid, caffeic acid and ferulic acid, involving phenylalanine ammonia lyase (PAL), hydroxylations, an *O*-methylation and finally a chain-shortening reaction. Formation of vanillin glucoside would also require the involvement of a UGT, although the point at which the glycosylation would take place remains elusive. Enzymes are known that can catalyse most of these reactions, for example, PAL<sup>20</sup>, cinnamic acid 4-hydroxylase, C4H<sup>21</sup>, *p*-coumaric acid 3-hydroxylase, C3H<sup>22</sup>, but is not clear whether CoA-derivatives are involved or whether the C3-hydroxylation step proceeds, for example, through quinate- and shikimate esters<sup>22</sup>. Caffeic acid could in principle be *O*-methylated<sup>23</sup> to form ferulic acid, the substrate for the hypothesized final chain-shortening reaction, and several OMTs are known from *V. planifolia*. Vanillin UGTs or genes that encode these enzymes from *V. planifolia* have not yet been reported and as stated above it remains to be demonstrated at which stage in the pathway glycosylation occurs.

The aim of the current study was to elucidate the vanillin biosynthesis pathway in *V. planifolia*. We have carried out biosynthetic studies with fresh vanilla pods using a number of different putative radiolabelled precursors. Supported by a combination of transcriptomic and proteomic approaches, we

identified a gene, *VpVAN* encoding a two-carbon chain-shortening enzyme converting ferulic acid and its glucoside directly into vanillin and its glucoside. *VpVAN* is produced in cells in the inner part of the vanilla pod. The substrate specificity of *VpVAN* was determined using a rabbit reticulocyte-coupled *in vitro* transcription/translation system and by heterologous expression of the gene in *Nicotiana benthamiana*, *Hordeum vulgare* and *Saccharomyces cerevisiae*. A gene sequence identical to that of *VpVAN* was previously reported to encode an enzyme designated 4-HBS catalysing a two-carbon chain-shortening of *p*-coumaric acid into 4-hydroxybenzaldehyde. We could not verify such a function in our studies.

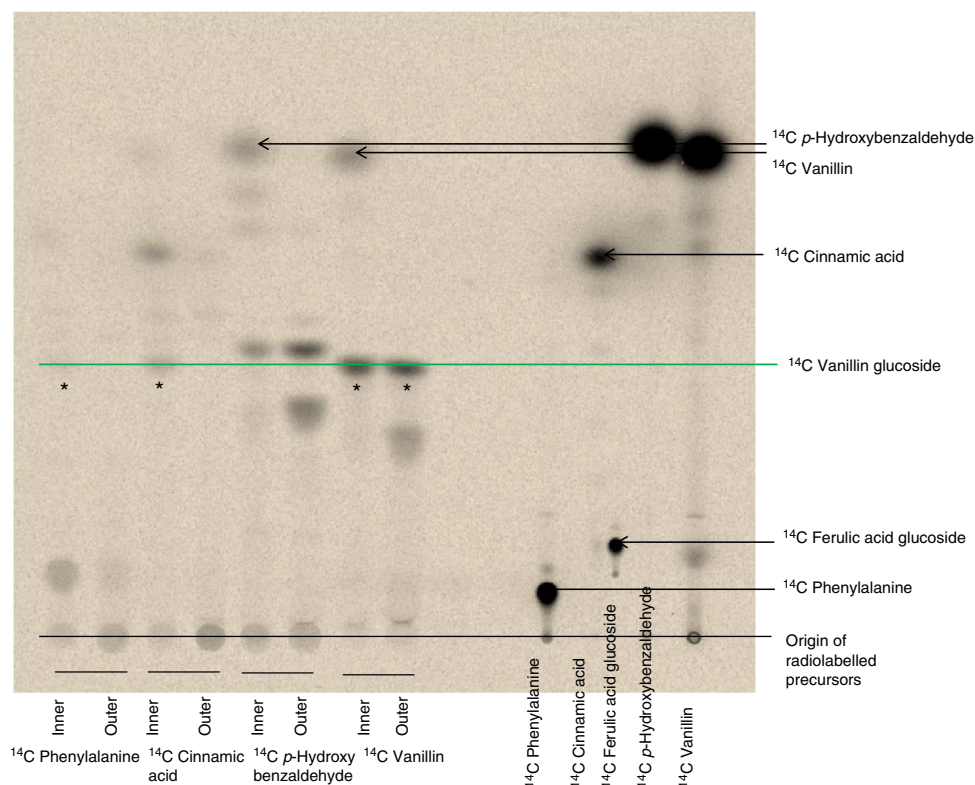
## Results

### Administration of putative [<sup>14</sup>C]- precursors to vanilla pods.

To examine whether the *p*-hydroxybenzaldehyde-based or the longer lignin precursor-based pathway is the most likely native vanillin glucoside biosynthetic pathway, [<sup>14</sup>C]-radiolabelled putative precursors ([<sup>14</sup>C]-phenylalanine, [<sup>14</sup>C]-cinnamic acid, [<sup>14</sup>C]-*p*-hydroxybenzaldehyde and [<sup>14</sup>C]-vanillin) were administered to sliced discs of fresh vanilla pods harvested 6 months after pollination. Vanillin glucoside is located in the inner part of the pod, that is, in the papillae and placental tissues, but completely absent from the epicarp, outer mesocarp area and seeds<sup>24</sup>. Accordingly, the experiments with administration of radiolabelled precursors were carried out separately with inner and outer parts of the pod discs using the outer parts of the pod as negative controls. Incubation with [<sup>14</sup>C]-phenylalanine and [<sup>14</sup>C]-cinnamic acid resulted in [<sup>14</sup>C]-vanillin glucoside formation in the tissue representing the inner part of the pod while administration of [<sup>14</sup>C]-*p*-hydroxybenzaldehyde induced [<sup>14</sup>C]-*p*-hydroxybenzaldehyde glucoside formation in both the inner and the outer part of the pod (Fig. 1; Supplementary Fig. 1). Incubation with [<sup>14</sup>C]-*p*-hydroxybenzaldehyde did not result in [<sup>14</sup>C]-vanillin glucoside formation. The radiolabelling studies confirmed that vanillin glucoside biosynthesis occurs only in the inner part of the pod and demonstrated that *p*-hydroxybenzaldehyde is not an intermediate in vanillin biosynthesis. The incorporation percentages observed varied depending upon the pod developmental stage, whereas the pattern of radiolabelled compounds observed following administration of each of the different labelled precursors at different developmental stages remained similar.

Administration of [<sup>14</sup>C]-vanillin resulted in formation of [<sup>14</sup>C]-vanillin glucoside both in the inner and outer part of the pod. Similarly [<sup>14</sup>C] *p*-hydroxybenzaldehyde administration resulted in [<sup>14</sup>C]-*p*-hydroxybenzaldehyde glucoside both in the inner and outer part of the pod, demonstrating the presence of a glucosyltransferase capable of glycosylating these precursors.

**Identification of candidate genes.** To identify genes and enzymes involved in vanillin glucoside biosynthesis in *V. planifolia*, a combination of transcriptomic and proteomic approaches was undertaken with an initial focus on candidates representing the five major enzyme families suggested from the literature to play a possible role in vanillin biosynthesis, namely PAL, cytochrome P450s (the monophenol monooxygenases C4H and C3H), OMTs, UGTs and the carbon chain-shortening enzyme, 4-HBS. The *V. planifolia* transcriptome was obtained from a 6-month-old vanilla pod from the island of La Réunion by 454 pyrosequencing. Approximately 40 UGTs, 15 OMTs, a CYP98A3 (C3H) and 4-HBS conreads were found in the transcriptome. To further assess the likelihood of involvement of each of these genes in vanillin biosynthesis, a targeted proteomic approach (proteomic mass finger printing) was carried out in parallel with the broad



**Figure 1 | Thin layer chromatography (TLC) analysis of the radiolabelled products formed following administration of different putative  $^{14}\text{C}$ -labelled precursors to tissue slices of fresh vanilla pods.** The  $^{14}\text{C}$ -labelled products formed following 48 h incubation were extracted into 70% MeOH, separated by TLC and monitored by autoradiography. The position of the radiolabelled precursors following chromatographic separation is shown on the right part of the thin layer. (\*) indicates formation of  $^{14}\text{C}$ -labelled vanillin glucoside from the administered precursor. The presence of ultraviolet-absorbing components in the vanilla extracts as well as the position of the putative substrates after chromatographic separation was also monitored following exposure to ultraviolet light (254 nm) (Supplementary Fig. 1).

transcriptome analysis using the biosynthetically active inner part of the pod as experimental tissue. On the basis of overlay of the pyrosequencing and proteomic data sets, we selected and cloned 1UGT (*VpUGT72U1*), 11OMTs, a *CYP98A3* orthologue (*VpCYP98A70*) and 4-HBS (Supplementary Table 1; Supplementary Data 1 and 2).

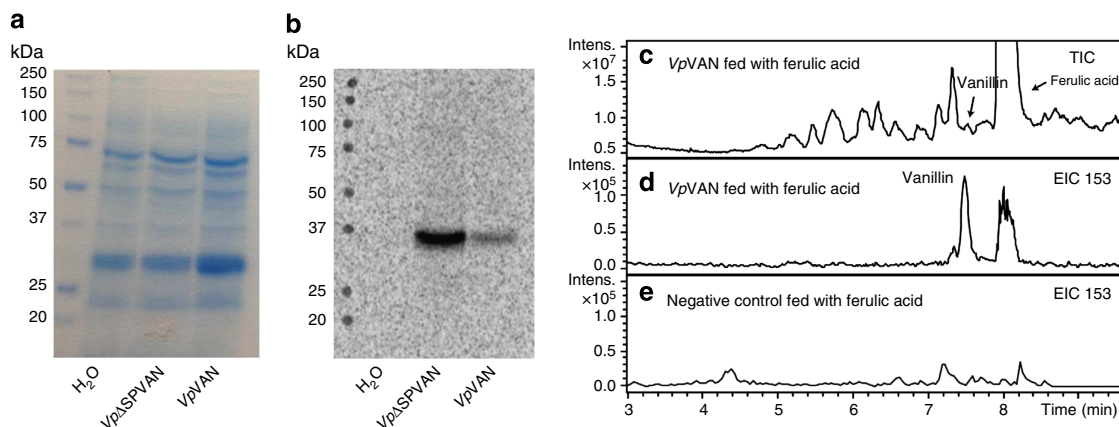
Although in the literature, the vanillin biosynthetic pathway has been suggested to be embedded within a metabolic grid, our initial *in vitro* studies with these gene candidates identified a gene encoding an enzyme converting ferulic acid glucoside and ferulic acid directly into vanillin glucoside and vanillin, respectively. This represents the first committed step in vanillin synthesis and demonstrates that vanillin formation in *V. planifolia* is catalysed by a single enzyme using a general substrate from phenylpropanoid metabolism. We designated the enzyme vanillin synthase and the gene *VpVAN* (gene sequence is given in Supplementary Fig. 2a). In a published patent application, the identical gene sequence had previously been assigned as encoding an enzyme converting *p*-coumaric acid into *p*-hydroxybenzaldehyde<sup>18</sup>. Accordingly the gene sequence was initially designated as encoding a *p*-hydroxybenzaldehyde synthase (4-HBS), an activity that we did not observe in our studies as reported below.

#### Vanillin synthase catalyses vanillin formation *in vitro*.

The catalytic activity of vanillin synthase using a range of different putative substrates was monitored by *in vitro* coupled transcription/translation assays. Vanillin synthase protein was obtained from its PCR-generated DNA in a coupled

transcription/translation assay with the inclusion of L-[ $^{35}\text{S}$ ]-methionine to provide easy monitoring of protein formation by SDS-polyacrylamide gel electrophoresis (PAGE) analysis (Fig. 2). The coupled assay produced a single radiolabelled protein band migrating with an apparent molecular mass of 36 kD in close agreement with the predicted mass of 39.15 kD for *VpVAN* (Fig. 2), thus enabling us to monitor the activity of the enzyme in an *in vitro* condition equalling a purified enzyme.

The substrate specificity of the *VpVAN* enzyme formed was investigated by incubation (1 h and 24 h) with 0.5–5 mM of *p*-coumaric acid, caffeic acid, ferulic acid, *p*-coumaric acid glucoside, caffeic acid glucoside, ferulic acid glucoside, feruloyl-CoA, *p*-coumaroyl-CoA and caffeoyl-CoA. Liquid chromatography-mass spectrometry (LC-MS) analyses demonstrated that *VpVAN* catalysed a direct chain shortening of ferulic acid and ferulic acid glucoside to vanillin and vanillin glucoside, respectively (Figs 2 and 3), whereas no activity was found using *p*-coumaric acid, caffeic acid and the glucosides of these as substrates. The conversion of ferulic acid and ferulic acid glucoside into vanillin and vanillin glucoside proceeded in the absence of added adenosine tri-phosphate (ATP) and nicotinamide adenine dinucleotide ( $\text{NAD}^+$ ). These cofactors would have been required if the chain shortening had occurred via  $\beta$ -oxidation of an activated CoA ester<sup>16</sup>. In a number of experiments, vanillin formation was observed using feruloyl-CoA as a substrate. In such experiments, the incubation mixture was found always to contain ferulic acid present as an impurity in the commercially provided feruloyl-CoA. Since ATP and  $\text{NAD}^+$  were not required as cofactors in these reactions, we conclude that vanillin synthase is



**Figure 2 | Direct coupled transcription/translation of the PCR-generated DNA for VpVAN and for VpVAN devoid of its 21 amino-acid-long ER-targeting signal peptide (VpΔSPVAN).** L-[<sup>35</sup>S]-methionine was included to specifically monitor the formation of *de novo* synthesized radiolabelled proteins by SDS-PAGE. The ability of VpVAN synthesized by *in vitro* transcription/translation to catalyse conversion of ferulic acid into vanillin was monitored by LC-MS using total and selected ion monitoring. **(a)** Proteins present in the *in vitro* transcription/translation assay visualized by Coomassie brilliant blue staining. **(b)** The [<sup>35</sup>S]-labelled VpΔSPVAN and VpVAN proteins formed from the two PCR products as visualized by autoradiography. In each transcription/translation experiment, a single radiolabelled protein band of the expected approximate mass was obtained. **(c)** Incubation of the transcription translation protein solution containing VpVAN with 5 mM of ferulic acid for 1 h in 2.5 mM dithiothreitol at 30 °C, total ion chromatogram (TIC) following LC-MS analysis; **(d)** EIC: 153: extracted ion chromatogram for *m/z* vanillin ( $M + H^+$ ) for specific detection of vanillin formation. **(e)** EIC 153: extracted ion chromatogram for *m/z* vanillin ( $M + H^+$ ) of a control experiment in which ferulic acid was administered to a transcription translation protein solution not expressing VpVAN. Intens., intensity.

not able to utilize feruloyl-CoA as substrate for vanillin formation by  $\beta$ -oxidation of the activated CoA ester<sup>16</sup>. Accordingly, we conclude that VpVAN is catalysing vanillin and vanillin glucoside formation from ferulic acid and ferulic acid glucoside in a coupled non-oxidative hydratase/lyase reaction.

A general amino-acid sequence identity search using GenBank showed that the VpVAN protein exhibits high sequence identity to cysteine proteinases. The highest amino-acid sequence identity (77%) was found to the *Elaeis guineensis* cysteine proteinase belonging to the aleurain class of cysteine proteinases (MEROPS—the peptidase database). Interestingly, alignments unequivocally demonstrated that the VpVAN sequence contained the three key active site residues required for proteinase activity<sup>25</sup>. Likewise the VpVAN amino-acid sequence contains the typical processing motif known from cysteine proteinases, indicating that the VpVAN protein when produced in the vanilla pod may be subject to a maturation cycle involving cleavage after the amino-acid residue at position 137 (DGV/LPVT). To test whether the activity of the VpVAN enzyme would be altered in the absence of the pro-peptide or if it changes the catalytic activity of the enzyme *in vitro*, we generated two truncated versions of VpVAN. The first truncated version was designed to lack the first 137 amino acids (vpΔ137van). A second truncated version of VpVAN was designed lacking the first 61 amino acids (vpΔ61van) to reflect processing at a second possible peptide cleavage site predicted by the ProP 1.0 Server. Hydropathy plot analysis indicated that the VpVAN sequence also contains an N-terminal endoplasmic reticulum (ER)-targeting signal peptide (first 21 amino-acid residues), and a truncated VpVAN with no ER-targeting signal peptide (vpΔspvan) was therefore also constructed and tested *in vitro* using the coupled transcription/translation assay. Neither the presence of the target sequence nor the presence of the pro-peptide has a significant influence on the activity of the VpVAN enzyme (Fig. 3).

**Cellular localization of vanillin synthase in *V. planifolia*.** The cellular localization of transcripts encoding vanillin synthase in the vanilla orchid was determined. The proteome analyses

showed that the VpVAN enzyme is localized in the inner part of the pod, the tissue actively synthesizing vanillin glucoside. In tube *in situ* PCR<sup>26</sup> was used to determine the cellular localization of transcripts encoding VpVAN (Fig. 4). The analyses were performed on tissue sections from vanilla pod discs using primers specific to VpVAN. Six-month-old pods were selected based on their high biosynthetic activity compared with younger pods. VpVAN transcripts were detected in the segment of the pod discs representing the inner part of the vanilla pod. High transcript levels were observed in single cells located a few cell layers from the epidermis facing the inner cavity of the pod (Fig. 4), supplementing the conclusion from the biosynthetic and proteomic studies that vanillin glucoside biosynthesis occurs in the inner part of the pod.

**The catalytic activity of VpVAN in yeast.** In parallel with the experiment to test the activity of VpVAN in *in vitro* transcription/translation assays, the catalytic activity of VpVAN was further investigated by heterologous expression in *S. cerevisiae* (Fig. 5) using a range of putative substrates: *p*-coumaric acid, caffeic acid, ferulic acid, *p*-coumaric acid glucoside, caffeic acid glucoside, ferulic acid glucoside and feruloyl-CoA, *p*-coumaroyl-CoA and caffeoyl-CoA. For the yeast experiment, we have used the yeast codon-optimized VpVAN (VpScVAN gene sequence; Supplementary Fig. 2b). The specificity of the VpScVAN enzyme was tested in yeast cells (Fig. 5) expressing vanillin synthase together with *Arabidopsis thaliana* UGT72E2 (*AtUGT72E2*) or together with *V. planifolia* UGT72U1 (*VpUGT72U1*). *VpUGT72U1* is a novel UGT selected based on the combined *V. planifolia* transcriptomic and proteomic study (Supplementary Tables 1 and 2; Supplementary Data 1 and 2). *VpUGT72U1* specifically glucosylates vanillin (Supplementary Fig. 3), whereas *AtUGT72E2* has an extensive substrate specificity and is able to catalyse glucosylation of vanillin as well as the phenolic hydroxyl group of ferulic acid, *p*-coumaric acid, caffeic acid, *p*-hydroxybenzaldehyde and protocatechuicaldehyde.

Yeast has previously been reported to efficiently reduce vanillin to vanillyl alcohol<sup>6</sup>. The current studies were carried out using



*S. cerevisiae* strain Y06460 in which alcohol dehydrogenase 6 (*ADH6*) is disrupted, because use of this strain circumvents reduction of the vanillin formed into vanillyl alcohol<sup>6</sup>. Genes were also integrated to simultaneously disrupt *EXG1* encoding an

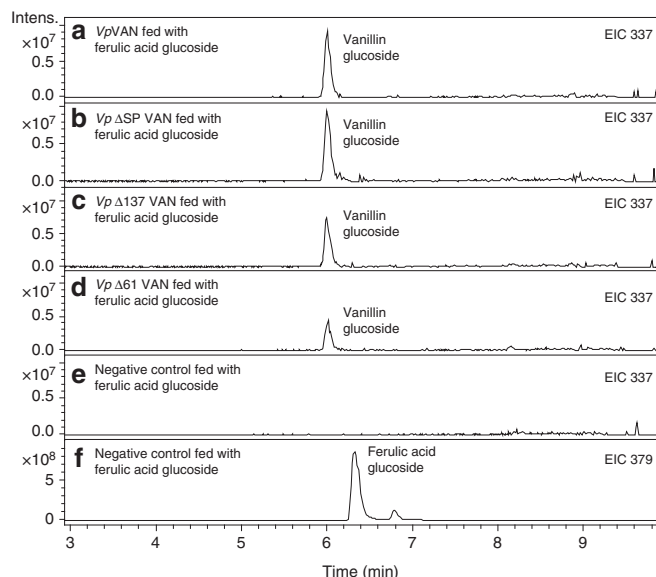
endogenous yeast exoglucosidase 1 (*EXG1*), which efficiently hydrolyses vanillin glucoside<sup>6</sup>. *VpVAN* was then stably integrated into the yeast chromosome either alone or together with *AtUGT72E2* or together with *VpUGT72U1*.

When *VpScVAN* was expressed alone in the yeast strain modified as described above in the presence of ferulic acid, no vanillin glucoside peak was observed in spite of the fact that the yeast endogenous exoglucosidase 1 (*EXG1*) had been knocked out. This demonstrated that the yeast does not contain a UGT able to glucosylate vanillin. Combined expression of *VpScVAN* and *VpUGT72U1* resulted in formation of vanillin glucoside when yeast was supplied with ferulic acid or ferulic acid glucoside. In addition, combined expression of *VpScVAN* and *AtUGT72E2* resulted in formation of vanillin glucoside when administered with ferulic acid or ferulic acid glucoside. These studies demonstrate that the vanillin synthase is able to catalyse carbon double-bond cleavage of both ferulic acid and ferulic acid glucoside (Fig. 5).

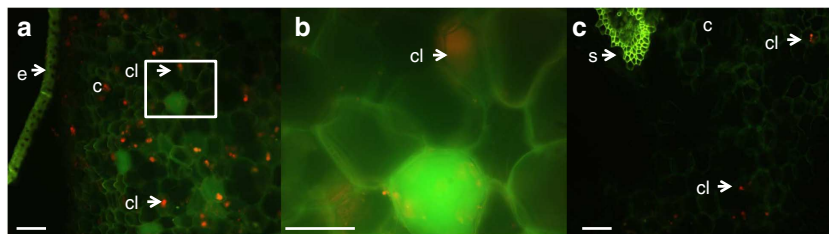
As previously mentioned, *VpVAN* was predicted to encode a protein with an ER-targeting signal peptide. Accordingly, an additional series of biosynthetic studies were carried out with yeast harbouring stably integrated *AtUGT72E2* together with either *VpVAN* or *VpScVAN* or truncated *VpVAN* with no ER-targeting signal peptide (*vpΔspvan*) or truncated *VpVAN* with no signal peptide and yeast codon optimized (*vpscΔspvan*). The yeast strains were incubated with putative substrates for 72 h and metabolite profiles determined by LC-MS. Formation of vanillin glucoside was observed with ferulic acid as substrate with *VpΔSpVAN* and *VpScΔSpVAN* (Supplementary Fig. 4). Thus *VpVAN* is catalytically active towards ferulic acid in the presence as well as absence of the ER-targeting signal peptide. Carbon chain shortening of caffeic acid and *p*-coumaric acid or glucosides of these was not observed with any of the modified versions of *VpVAN*.

Concomitant with the conversion of ferulic acid and its glucoside into vanillin and vanillin glucoside, yeasts are able to metabolize ferulic acid into 4-vinylguaiacol. The latter conversions are highly significant (Fig. 5). Two enzymes in *S. cerevisiae* are known to be responsible for the conversion of ferulic acid to 4-vinylguaiacol. These are phenylacrylate decarboxylase (*PAD1*) and ferulate decarboxylase (*FDC1*)<sup>27–29</sup>. Increased levels of *VpVAN*-based vanillin production in yeast would thus be envisioned following disruption or downregulation of the two genes encoding *PAD1* and *FDC1*.

Molasses are obtained as by-products in the production of sugar from sugar beets, sugar cane or sorghum, and these molasses contain ferulic acid<sup>30,31</sup>. To examine whether such

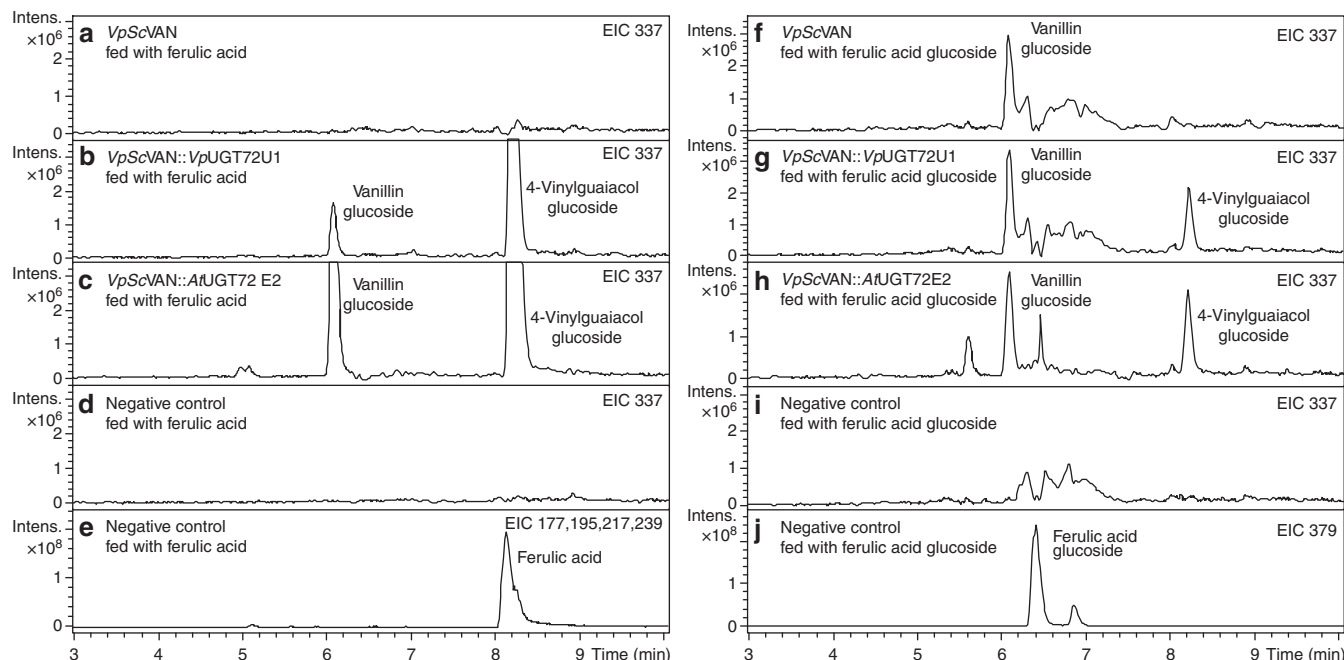


**Figure 3 | The ability of enzymes synthesized by *in vitro* transcription/translation to catalyse conversion of ferulic acid glucoside into vanillin glucoside.** The experiment shown in (a–d) involved incubation of the transcription/translation protein solutions with *VpVAN*, *VpΔSPVAN*, *VpΔ137VAN* and *VpΔ61VAN* with 1 mM of ferulic acid glucoside for 1 h in 1 mM dithiothreitol at 30 °C. (a) EIC 337: extracted ion chromatogram *m/z* vanillin glucoside ( $M + Na^+$ ) obtained following incubation of *VpVAN* with ferulic acid glucoside (b) EIC 337: extracted ion chromatogram *m/z* vanillin glucoside ( $M + Na^+$ ) obtained following incubation of *VpΔSPVAN* with ferulic acid glucoside (c) EIC 337: extracted ion chromatogram *m/z* vanillin glucoside ( $M + Na^+$ ) obtained following incubation of *VpΔ137VAN* with ferulic acid glucoside. (d) EIC 337: extracted ion chromatogram *m/z* vanillin glucoside ( $M + Na^+$ ) obtained following incubation of *VpΔ61VAN* with ferulic acid glucoside. (e) EIC 337: extracted ion chromatogram *m/z* vanillin glucoside ( $M + Na^+$ ) obtained following incubation of a control transcription translation/ protein solution devoid of any protein of interest with ferulic acid glucoside. (f) EIC 379: extracted ion chromatogram *m/z* ferulic acid glucoside ( $M + Na^+$ ) obtained following incubation of a control transcription translation/ protein solution devoid of any protein of interest with ferulic acid glucoside. Intens., intensity.



**Figure 4 | Tissue localization of the expression of *VpVAN* in transverse sections of a 6-month-old vanilla pod determined by *in tube in situ* PCR.**

Transcripts of *VpVAN* were detected in specific cells (Panels a and b) using FITC-conjugated antibodies recognizing digoxigenin (DIG) incorporated in the specific PCR products representing the *VpVAN* transcript. Higher magnification of the selected area in Panel a is shown in Panel b. No transcripts of *VpVAN* were detected in the controls in the absence of specific primers for *VpVAN* (Panel c). The fluorescence detected in Panel c represents unspecific binding of the FITC-conjugated antibodies recognizing DIG to cell walls and supporting fibre cells surrounding the vascular systems. In all panels the chloroplasts are visible owing to their auto fluorescence at the used filter settings. The sections were analysed with a Leica FI/RH filter with excitation filters: BP490/15; 560/25 and emission filters: BP525/20; 605/30. e, epidermis; c, cortex; cl, chloroplast; s, supporting fibre tissue; Scale bar, 100 μm.



**Figure 5 | The ability of VpScVAN expressed in an adapted yeast strain to metabolize ferulic acid and ferulic acid glucoside into vanillin and vanillin glucoside, respectively.** (a) Expression of VpScVAN does not result in formation of vanillin glucoside. Extracted ion chromatogram (EIC) 337:  $m/z$  vanillin glucoside ( $M + Na^+$ ). (b) Co-expression of VpScVAN and VpUGT72U1 results in formation of low amounts of vanillin glucoside. EIC 337:  $m/z$  vanillin glucoside ( $M + Na^+$ ). (c) Co-expression of VpScVAN and AtUGT72E2 results in formation of larger amounts of vanillin glucoside. EIC 337:  $m/z$  vanillin glucoside ( $M + Na^+$ ). (d) Negative control demonstrating that administration of ferulic acid to the adapted yeast strain does not result in vanillin glucoside formation. EIC 337:  $m/z$  vanillin glucoside ( $M + Na^+$ ). (e) Negative control monitoring the level of ferulic acid substrate administered. EIC 177, 195, 217, 239:  $m/z$  ferulic acid ( $M - OH = 177$ ,  $M + H^+ = 195$ ,  $M + Na^+ = 217$  and  $M + 2Na^+ = 239$ ). Note that the EIC trace for vanillin glucoside ( $M + Na^+ = 337$ ) also monitors the presence of 4-vinylguaiaicol glucoside ( $M + Na^+ = 335$ ) because this compound is present in such large amounts that the  $M + 2$  mass representing the natural isotope distribution is also recorded. (f) Expression of VpScVAN results in formation of vanillin glucoside. EIC 337:  $m/z$  vanillin glucoside ( $M + Na^+$ ). (g) Co-expression of VpScVAN and VpUGT72U1 does not augment vanillin glucoside formation. EIC 337:  $m/z$  vanillin glucoside ( $M + Na^+$ ). (h) Co-expression of VpScVAN and AtUGT72E2 does not augment vanillin glucoside formation. EIC 337:  $m/z$  vanillin glucoside ( $M + Na^+$ ). (i) Empty vector control demonstrating that administration of ferulic acid glucoside to the adapted yeast strain does not result in vanillin glucoside formation. EIC 337:  $m/z$  vanillin glucoside ( $M + Na^+$ ). (j) Empty vector control monitoring the level of ferulic acid glucoside substrate administered. EIC 379:  $m/z$  ferulic acid glucoside ( $M + Na^+$ ). Note that the EIC trace for vanillin glucoside ( $M + Na^+ = 337$ ) also monitors the presence of 4-vinylguaiaicol glucoside ( $M + Na^+ = 335$ ) because this compound is present in such large amounts that the  $M + 2$  mass representing the natural isotope distribution is also recorded. Intens., intensity.

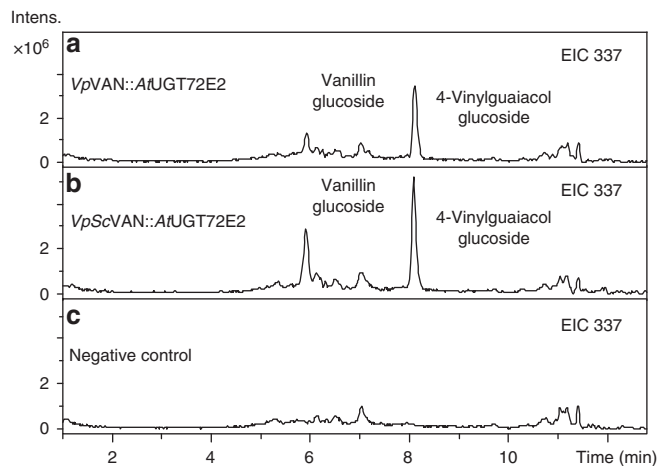
material could be used for vanillin glucoside production, yeasts expressing VpVAN and VpScVAN as well as AtUGT72E2 were grown in molasses-based growth medium using sugar beet as the source for the molasses. Vanillin glucoside formation was observed with both versions of VpVAN, highlighting the potential of this enzyme for vanillin glucoside production based on inexpensive waste materials (Fig. 6).

**Establishing vanillin synthesis in tobacco and barley.** The catalytic activity of vanillin synthase in the presence and absence of a putative pro-peptide *in vivo* was analysed following transient expression in tobacco and stable expression in barley. The *in vivo* biological activity of VpVAN (including the ER-targeting signal peptide) was assessed by transient expression in leaves of *N. benthamiana* in the absence of any exogenously added substrates. Gene constructs were transferred to *Agrobacterium tumefaciens* and co-infiltrated with an *A. tumefaciens* strain harbouring the *p19* gene-silencing suppressor. LC-MS profiling showed VpVAN-dependent formation of vanillyl alcohol glucoside (Fig. 7). The vanillyl alcohol glucoside arises by reduction of vanillin by an alcohol dehydrogenase (E.C.1.1.1.1) and subsequent glucosylation of the primary alcohol group of vanillyl alcohol, as was previously observed in cell cultures of

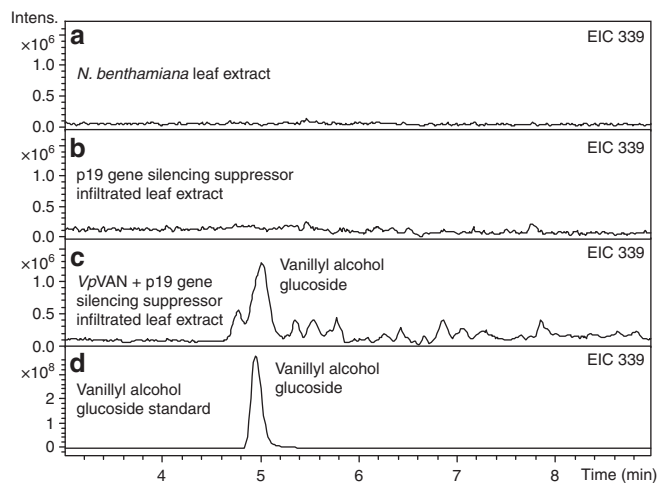
*N. plumbaginifolia*<sup>32</sup> and yeast<sup>6</sup>. Biotechnological production of vanillin glucoside in plants other than *Vanilla* sp. by introduction of VpVAN thus requires co-expression of a UGT that effectively glucosylates the free vanillin formed into the corresponding glucoside before its reduction into vanillyl alcohol.

Transient expression of *vpΔ137van* and *vpΔ61van* was also included in this study to investigate the importance of secondary processing of VpVAN for its *in vivo* activity. Introduction of each of these constructs encoding different truncated forms of VpVAN was found to result in vanillyl alcohol glucoside production in similar amounts as observed with VpVAN (Fig. 8a).

As previously mentioned, the VpVAN sequence showed high sequence identity to proteins belonging to the family cysteine proteinases. We identified a protein belonging to the family of cysteine proteinases in tobacco in which the amino-acid sequence identity to VpVAN was 71% (*N. benthamiana* cysteine proteinase gene sequence; supplementary Fig. 2c). In order to investigate the possibility to produce a nascent protein more amenable to proper targeting and processing by the endogenous tobacco machinery, a gene construct was made where the VpVAN ER-targeting signal peptide and pro-peptide protease cleavage site were replaced with the putative signal peptide and the putative pro-peptide protease cleavage site from the tobacco cysteine protease (*vpnbΔspΔ137van* gene sequence; Supplementary Fig. 2d).



**Figure 6 | Vanillin glucoside formation in a molasses-based growth medium following incubation with yeast expressing VpVAN or yeast codon-optimized VpScVAN as well as AtUGT72E2 stably integrated into the yeast genome.** Both wild-type (panel a, EIC 337:  $m/z$  vanillin glucoside ( $M + Na^+$ )) and yeast codon-optimized versions (panel b, EIC 337:  $m/z$  vanillin glucoside ( $M + Na^+$ )) of VpVAN were stably integrated into the yeast chromosome together with AtUGT72E2. Panel c is the wild type yeast strain (EIC 337:  $m/z$  vanillin glucoside ( $M + Na^+$ )). The yeast strains were grown in Delft medium supplemented with 8% molasses and the production of vanillin glucoside was monitored by LC-MS and extracted ion chromatography. Intens., intensity.



**Figure 7 | The biological activity of VpVAN assessed by a transient expression in *Nicotiana benthamiana* as analysed by LC-MS and illustrated by extracted ion chromatogram (EIC).** (a,b) Vanillyl alcohol glucoside is neither present in non-transformed *N. benthamiana* leaf extract nor in p19 gene-silencing suppressor infiltrated leaf extract (EIC 339:  $m/z$  vanillyl alcohol glucoside ( $M + Na^+$ )). (c) VpVAN was transferred to *Agrobacterium tumefaciens* and co-infiltrated with an *A. tumefaciens* strain harbouring the p19 gene-silencing suppressor in *N. benthamiana* leaves. Four days after inoculation, the infiltrated tobacco leaves were harvested and subjected to metabolite profiling. EIC 339:  $m/z$  vanillyl alcohol glucoside ( $M + Na^+$ ). (d) Vanillyl alcohol glucoside standard. EIC 339:  $m/z$  vanillyl alcohol glucoside ( $M + Na^+$ ). Intens., intensity.

The resulting construct *vpnbΔspΔ137van* was transferred to *A. tumefaciens* and transiently expressed in tobacco following infiltration. LC-MS profiling and extracted ion chromatography showed that modification of the VpVAN sequence by insertion of the tobacco target sequence and pro-peptide protease cleavage

site resulted in a several fold higher production of vanillyl alcohol glucoside in comparison with the amounts obtained from the VpVAN sequence (Fig. 8e).

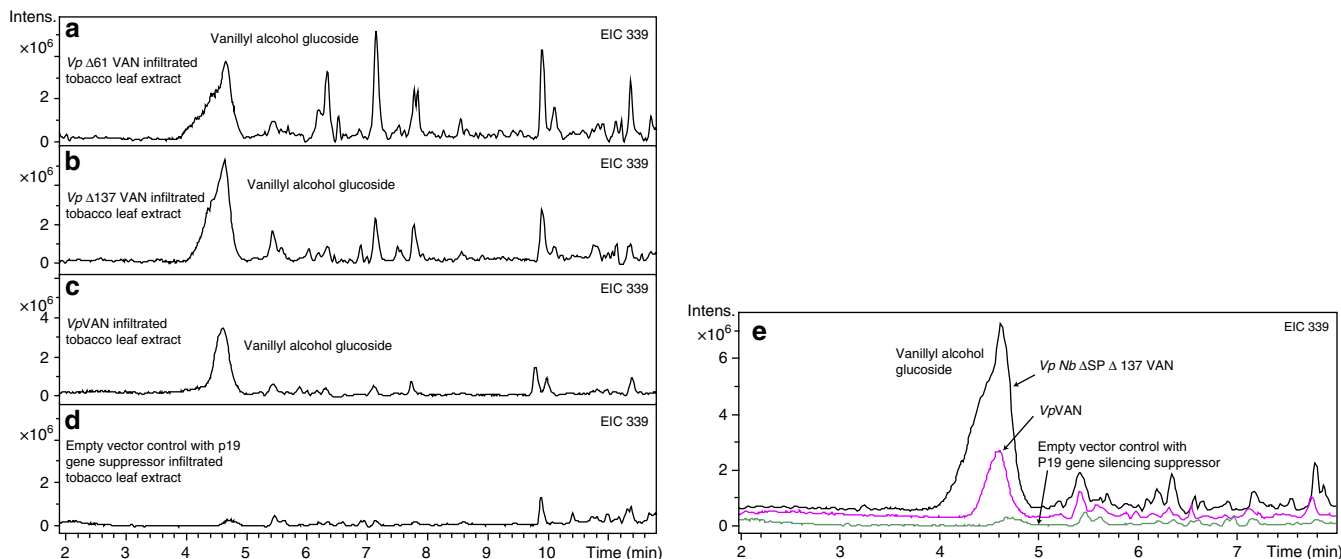
Plants belonging to the *Poaceae* family are known to accumulate higher amounts of ferulic acid and ferulic acid glucoside compared with other plant families<sup>33</sup>. It was therefore of interest to investigate the effects of stable *in vivo* expression of VpVAN in barley. In one series of transformations, the VpVAN gene sequence including the part encoding the ER-targeting signal peptide was codon optimized for barley (VpHvVAN gene sequence; Supplementary Fig. 2e). In a second series of transformations, the VpHvVAN gene sequence was additionally modified to encode a D-hordein signal peptide as a replacement for the original vanilla ER-targeting signal peptide (*vpHvΔspvan*). A constitutive ubiquitin promoter was used to drive the expression of both genes. Leaf samples from successfully transformed plants were collected 6–8 weeks after transfer of plantlets to the greenhouse and metabolic profiling was carried out by LC-MS. Barley plants transformed with *vpHvΔspvan* were found to accumulate vanillyl alcohol glucoside in significantly higher levels than control plants (Supplementary Fig. 5).

**Glechoma hederacea contains a vanillin synthase homologue.** A study of volatile constituents released from *G. hederacea* (ground ivy) belonging to the Lamiaceae family had shown that leaves of this plant release traces of vanillin<sup>2</sup>. RNA was isolated from the leaves. Transcriptome analysis identified an RNA sequence encoding a protein sequence showing 71% amino-acid sequence identity to VpVAN. To investigate whether the ability to produce vanillin could be assigned to the expression of this gene, the gene was transiently expressed in tobacco. Analysis of the tobacco leaf extracts demonstrated that expression of the gene resulted in accumulation of vanillyl alcohol glucoside. Thus the gene encodes a protein with similar functional properties as VpVAN. Accordingly the gene was assigned as *GhVAN* (Fig. 9) (*GhVAN* sequence: Supplementary Fig. 2f).

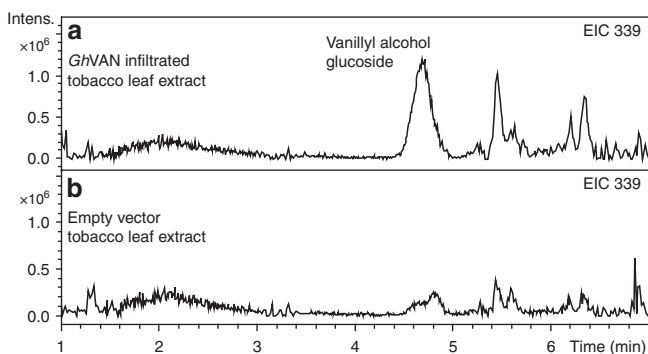
## Discussion

Numerous studies of the formation of vanillin and its glucoside have been carried out in the vanilla orchid *V. planifolia*, but no consensus biosynthetic pathway has emerged and specific enzymes involved in vanillin glucoside biosynthesis have not been conclusively demonstrated. In the current study, we have shown that the *de novo* biosynthesis of vanillin in the orchid *V. planifolia* and in *G. hederacea* (ground ivy), which belongs to the Lamiaceae family, is catalysed by a single enzyme, vanillin synthase that catalyses the two-carbon cleavage of ferulic acid and its glucoside to produce vanillin and vanillin glucoside, respectively (Fig. 10a). This conclusion was supported by biosynthetic experiments, which demonstrated that administration of the radiolabelled ferulic acid precursors phenylalanine and cinnamic acid to tissue slices of developing *V. planifolia* pods resulted in the formation of radiolabelled vanillin glucoside. These data combined with proteomic analysis demonstrated that vanillin biosynthesis takes place only in the inner part of the pod. This result is in accordance with previous observations<sup>24</sup>. Our *in situ* studies indicate that the vanillin synthase transcript and protein co-occur in the inner part of the pod corroborating the *in vivo* localization of the vanillin glucoside biosynthetic pathway.

Owing to the presence of high concentrations of 4-hydroxybenzaldehyde glucoside in mature pods and its structural similarity to vanillin glucoside, 4-hydroxybenzaldehyde or its glucoside has been proposed to be a precursor in the biosynthesis of vanillin and vanillin glucoside<sup>19</sup>. The huge pool of 4-hydroxybenzaldehyde glucoside found in mature fresh pods



**Figure 8 | The biological activity of VpVAN and modifications thereof assessed by a transient expression in *Nicotiana benthamiana*.** (a) Vanillyl alcohol glucoside is present in the extract of leaves of *N. benthamiana* expressing *vpΔ61van*. EIC 339: *m/z* vanillyl alcohol glucoside ( $M + Na^+$ ); (b) Vanillyl alcohol glucoside is present in the extract of leaves of *N. benthamiana* expressing *vpΔ137van*. EIC 339: *m/z* vanillyl alcohol glucoside ( $M + Na^+$ ). (c) Vanillyl alcohol glucoside is present in the extract of leaves of *N. benthamiana* expressing *VpVAN*. EIC 339: *m/z* vanillyl alcohol glucoside ( $M + Na^+$ ). (d) Control using an empty vector construct harbouring the *p19* gene-silencing suppressor, EIC 339: *m/z* vanillyl alcohol glucoside ( $M + Na^+$ ). Modifications of the *VpVAN* sequence have resulted in similar production of vanillyl alcohol glucoside in comparison with the amounts obtained from the *VpVAN* sequence. (e) *vpnbΔspΔ137van* was transferred to *A. tumefaciens* and transiently expressed in tobacco. Four days after inoculation, the infiltrated tobacco leaves were harvested and subjected to metabolite profiling. EIC 339: *m/z* vanillyl alcohol glucoside ( $M + Na^+$ ). Modification of the *VpVAN* sequence by insertion of the tobacco target sequence and pro-peptide protease cleavage site resulted in enhanced production of vanillyl alcohol glucoside in comparison with the amounts obtained from the *VpVAN* sequence. Intens., intensity.



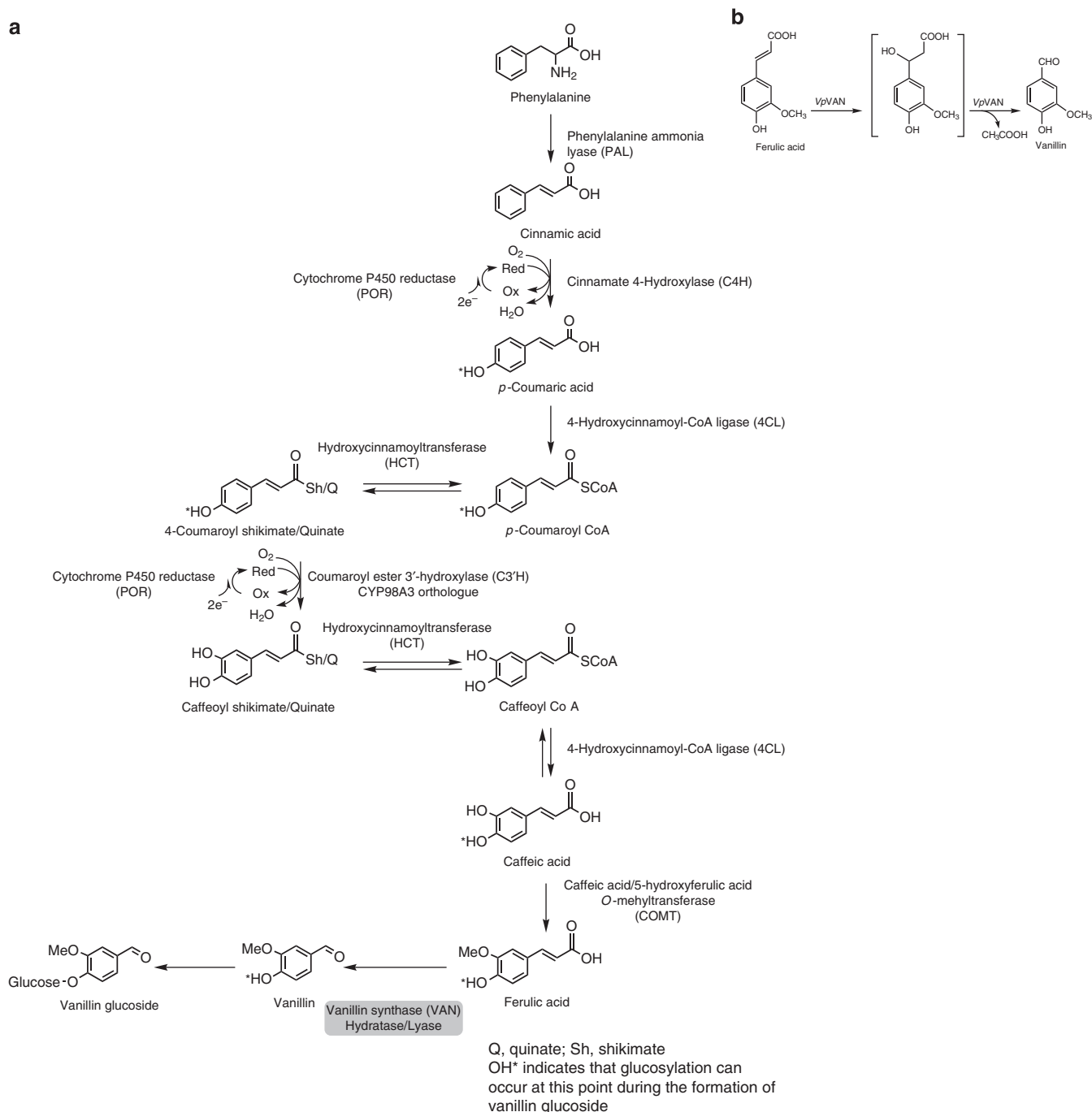
**Figure 9 | The biological activity of GhVAN assessed by a transient expression in *Nicotiana benthamiana*.** (a) Vanillyl alcohol glucoside is present in the extract of leaves of *N. benthamiana* expressing *GhVAN*. EIC 339: *m/z* vanillyl alcohol glucoside ( $M + Na^+$ ). (b) Control using an empty vector construct harbouring the *p19* gene-silencing suppressor. (EIC 339: *m/z* vanillyl alcohol glucoside ( $M + Na^+$ ). Intens., intensity.

could either represent accumulation of an excess of 4-hydroxybenzaldehyde glucoside that is not yet converted into vanillin or reflect a separate function of 4-hydroxybenzaldehyde *per se*. In the current study, we have shown that administration of [<sup>14</sup>C] 4-hydroxybenzaldehyde to the inner part of the pod does not result in radiolabelling of vanillin glucoside under conditions where vanillin glucoside is known to be formed. The *in planta* biosynthetic routes to 4-hydroxybenzaldehyde and benzoic acid in other plant species also remain partly unresolved<sup>17,34</sup>. In the developing vanilla pod, the phenylalanine-derived phenylpropanoids such as *p*-coumaric acid, ferulic acid and sinapic

acid may be directed towards formation of lignin monomers. The vanillin glucoside concentration increases with the age of the pod after pollination. The vanilla pod achieves its full-grown pod size about 3 months after pollination and is mature when about 10 months old. If left on the plant, the pod begins to split from the end, exposing its seeds. In the mature state, the pod only has a few requirements for *de novo* synthesis of cell wall components and the plant may therefore shift the flux of phenylpropanoid precursors from lignin biosynthesis to synthesis of vanillin glucoside and 4-hydroxybenzaldehyde glucoside in order to improve its potential for chemical defense of the maturing pod towards herbivores and pests. Vanillin and 4-hydroxybenzaldehyde exhibit anti-microbial properties. The metabolic changes in the course of pod ontogeny may thus serve to balance optimal pod development and defense.

The gene sequence that we have identified as encoding a vanillin synthase had previously been proposed to encode an enzyme, *p*-hydroxybenzaldehyde synthase (4-HBS), catalysing the conversion of *p*-coumaric acid into *p*-hydroxybenzaldehyde<sup>19</sup>. We tested the catalytic properties of the enzyme encoded by the gene sequence in coupled *in vitro* transcription/translation assays, following stable expression in yeast and following transient expression in tobacco and stable expression in barley. In the *in vitro* coupled transcription/translation system a range of putative substrates was provided including *p*-coumaric acid. In the transient and stable expression systems used, product formation was dependent on the availability of an endogenously produced substrate. In none of these experimental systems did we observe an ability of the *VpVAN* enzyme to catalyse the conversion of *p*-coumaric acid into *p*-hydroxybenzaldehyde or *p*-hydroxybenzalcohol glucoside (*in planta*). This was monitored by LC-MS analyses and extracted ion monitoring. In all cases, the presence of a free or glycosylated hydroxyl group at the 4th





**Figure 10 | The *de novo* biosynthesis of vanillin is mediated by a single enzyme VpVAN in the pods of *V. planifolia*.** (a) The *de novo* biosynthesis of vanillin is mediated by a single enzyme, namely *V. planifolia* vanillin synthase (VpVAN), which catalyses the two-carbon cleavage of ferulic acid and its glucoside to produce vanillin and vanillin glucoside, respectively. (b) The conversion of ferulic acid to vanillin is catalysed by VpVAN and is envisioned to proceed sequentially by two partial reactions composed of an initial hydration addition reaction followed by a retro-aldol elimination reaction. Intens., intensity.

position of the phenolic ring in combination with the presence of a methoxy group at the 3rd position was required for VpVAN to exert activity. Experiments to measure the enzyme activity of the VpVAN enzyme in protein extracts from the *V. planifolia* pod were not successful because the high amounts of endogenously produced vanillin glucoside and *p*-hydroxybenzaldehyde present prevented detection of minute additional amounts of product possibly formed following supplementation of precursors. Neither column chromatography nor dialysis for several days

was sufficient to lower the amount of endogenous vanillin glucoside present to an acceptable level. In the study of Podstolski *et al.*<sup>19</sup>, different interconvertible isoforms of 4-HBS were partly purified from embryo cell cultures of *V. planifolia*. One of those isoforms, impurities present or residual amounts of endogenously bound *p*-hydroxybenzaldehyde or a simultaneously reported spontaneous background reaction may have given rise to the observed *p*-hydroxybenzaldehyde formation.

The conversion of ferulic acid and its glucoside into vanillin and the corresponding glucoside is envisioned to proceed sequentially by two partial reactions composed of an initial hydration addition reaction followed by a retro-aldol elimination reaction (Fig. 10b). The initial reaction consists in the addition of water to the double bond. The  $\beta$ -hydroxy carboxylic acid formed then undergoes a well-known retro-aldol elimination reaction, which results in the formation of vanillin and acetic acid in stoichiometric amounts. Since our studies excluded the requirement for any cofactors, this remains the only plausible reaction mechanism, although we did not carry out assays to determine the acetate release.

This reaction mechanism has been demonstrated in some bacteria for the bioconversion of the CoA thioester of ferulic acid to vanillin, for example, in cultures of *Pseudomonas fluorescens* by the enzyme 4-hydroxycinnamoyl-CoA hydratase/lyase (HCHL). HCHL catalyses the degradation of a range of 4-hydroxycinnamic acid CoA thioesters including ferulic acid-CoA<sup>35,36</sup>. This bacterial enzyme is a member of the low sequence similarity hydratase/isomerase superfamily of enzymes also referred to as the crotonase superfamily. Enzymes belonging to this family are known to catalyse highly divergent types of reactions including hydratase/lyase reactions, and the specific function of individual family members cannot easily be deduced solely from their amino-acid sequence because the amino-acid residues specifying the activity are scattered throughout the entire protein sequence<sup>37</sup>. The intermediates in these reactions are usually thioester enolate anions stabilized by a conserved oxyanion hole through hydrogen bonds<sup>38</sup>. Two Glu residues serve as acid/base catalysts for the reaction, although in some members of the crotonase family the second Glu residue is absent<sup>38</sup>. A sequence alignment of VpVAN with the HCHL sequence from *P. fluorescens* shows insignificant sequence similarity. Of the 55 conserved amino-acid residues scattered over the entire protein sequence in selected bacterial sequences belonging to the crotonase superfamily (see Achterholt *et al.*<sup>39</sup>, Fig. 5), only 11 were similarly positioned in VpVAN. The alignment with VpVAN identifies the position of the two consensus sequences involved in stabilizing the oxyanion hole in the crotonase superfamily. These show low sequence identity YGSEEE (residues 67–72) and QGI (residues 147–149) to the consensus sequences found in crotonases<sup>38</sup>.

A general sequence identity search using GenBank showed that the VpVAN protein sequence has a much higher sequence identity to cysteine proteinases. Cysteine proteinases are expressed as a pre-protein with an N-terminal ER-targeting signal peptide being part of a pro-peptide domain containing 130–160 residues<sup>40</sup>. In the mature protein, the pro-peptide sequence is removed either with the aid of a processing enzyme or auto-catalytically<sup>41</sup>. Autocatalytic cleavage would have resulted in the formation of a protein with a mass of 23.89 kD. The *in vitro* transcription/translation experiments in which the VpVAN protein formed was labelled with <sup>35</sup>S methionine showed no evidence of autocatalytic processing (Fig. 2b) indicating that removal of the pro-peptide requires the action of a separate processing enzyme. Residues that are known to be conserved among different cysteine proteinases because they form part of the active site were also found to be conserved in VpVAN. These include Q156, C162, N301 and NSW322–24 (ref. 42). Likewise the six cysteine residues known to be involved in disulphide bridge formation in cysteine proteinases are conserved in VpVAN: C159/C202 C193/C235, C293/C343 (ref. 42). The non-contiguous ERFNIN signature (E  $\times$  3R  $\times$  3F  $\times$  3N  $\times$  3I/V  $\times$  3N) found in the pro-peptide of some groups of papain-like cysteine proteinases is also present in VpVAN (E72, R76, F80, N83, I87 and N91). The GC  $\times$  GG domain known from papaine-like cysteine proteinases

is also conserved in VpVAN (residues 201–205; Supplementary Fig. 2g). A putative sumoylation site FKME is located near the C-terminal end of VpVAN (residues 334–337 Supplementary Fig. 2g). SUMO modification of a tomato cysteine proteinase targeted the cysteinase to the nucleus where it activated a gene in the ethylene biosynthetic pathway<sup>43</sup>. This may imply a role of VpVAN in vanilla pod senescence. The pro-peptide released may act as an inhibitor of plant pests<sup>44,45</sup>. The catalytic mechanism of cysteine proteinases involves formation of a tetrahedral transition state composed of an oxyanion hole stabilized by hydrogen bonds. In the processed mature cysteine proteinase papain, the backbone amide of the catalytic C25 residue and the side chain amide of Q19 provide the hydrogen bonds<sup>46</sup>. These residues correspond to residues C162 and Q156 in VpVAN. The establishment of a transition state composed of an oxyanion hole stabilized by hydrogen bonds is thus a common feature of those two enzyme classes to which VpVAN shows sequence homology. We therefore propose that formation of a similar oxyanion transition state constitutes the initial step in the VpVAN catalysed conversion of ferulic acid and its glucoside into vanillin and vanillin glucoside, respectively. The oxyanion intermediate would facilitate hydration and constitute the intermediate that by a retro-aldol elimination reaction affords C–C cleavage of the propanoid side chain as required for vanillin formation.

Two putative protease cleavage sites in VpVAN were identified after residue 61(RFAR/RYGK) and residue 137 (VDGV/LPVT). The N-terminal pro-peptide sequence in VpVAN as well as in plant cysteine proteinases may be envisioned to serve different functions. The pro-peptide sequence may control proper intracellular targeting, may promote proper folding of the mature enzyme and may also serve to maintain the enzyme in an inactive form in the cell to balance its function according to physiological demands. In our studies, we have observed that VpVAN is also active in the presence of the pro-peptide sequence (Figs 3, 5a,b and 6 and so on) documenting that the presence of the pro-peptide does not severely inhibit VpVAN activity. However, when the VpVAN pro-peptide was replaced with the ER-targeting putative signal peptide and the putative pro-peptide protease cleavage site from the tobacco cysteine protease and transiently expressed in tobacco, higher levels of vanillin alcohol glucoside were obtained compared with parallel experiments with VpVAN. It is possible that the presence of the tobacco ER-targeting signal peptide and the pro-peptide protease cleavage site from the tobacco is able to target VpVAN to the correct cell compartment ensuring correct folding and post-translation modification of the protein, resulting in enhanced enzyme activity.

The identification of a hydratase/lyase type enzyme as being a vanillin synthase offers new opportunities for the *Vanilla* pod-based industries. The accumulation of vanillin glucoside in the capsules of cultivated vines in response to environmental challenges may now be assessed at the molecular level. Likewise, the basis for development of genetic markers for the selection of vanilla orchid varieties with improved aromatic properties has now been laid down. Vanillin produced biologically is termed 'natural' vanillin and has a high economic value compared with chemically synthesized vanillin. Likewise, in the transition towards a bio-based economy, it is important to develop sustainable production systems to replace those currently based on fossil fuels. The demonstration that a single enzyme in the vanilla pod catalyses the conversion of ferulic acid and ferulic acid glucoside into vanillin and vanillin glucoside provides several options for biotechnological applications. As demonstrated in the current study, molasses may be used for vanillin production based on their ferulic acid content and following supplementation with yeast expressing vanillin synthase, but are devoid of ferulate

decarboxylase activity. Ferulic acid is a key intermediate in lignin monomer formation in plants, so stable expression of *VpVAN* and, for example, *AtUGT72E2* in plants would be expected to result in vanillin glucoside formation in varying amounts. In cured *Vanilla* pods, only partial hydrolysis of the vanillin glucoside originally present has occurred offering a slow release aroma effect when residual amounts of the glucoside are hydrolysed by microbial enzymes in the mouth saliva following human ingestion. In pig production farms, addition of vanillin to the pig feed has been shown to increase fodder uptake by the pigs and their growth rate<sup>47</sup>. If so desired, transgenic plants with high vanillin synthase activity may be used as production sources for vanillin glucoside. Alternatively, yeast and other microorganisms may be used as production platforms following stable integration of genes encoding for enzyme conversion of phenylalanine into ferulic acid, vanillin synthase and a vanillin glucosyltransferase.

## Methods

**Plant material.** Healthy branches of *V. planifolia* carrying foliage and green vanilla pods were harvested at La Réunion 3 and 6 months after pollination and were shipped by courier carrier to Denmark while maintaining high-humidity conditions. Pod discs stabilized in RNAlater (Qiagen) were also imported from the biological resource center, VATEL, CIRAD, Saint-Pierre, La Réunion, France.

*N. benthamiana* plants (3 weeks old) were used for the transient expression of *VpVAN*.

The spring barley cultivar Golden Promise was grown in growth cabinets at a day/night temperature regime of 15/10 °C with a 16-h light period (light intensity: 350  $\mu\text{E m}^{-2} \text{s}^{-1}$ ). Immature embryos were isolated 12–14 days after pollination and used for Agrobacterium-mediated production of stably transformed barley plants.

**Strains and growth media.** *Escherichia coli* DH5 $\alpha$  (*endA1*, *hsdR17*, *gyrA96*, *thi-1*, *relA1*, *supE44*, *recA1*,  $\Delta$ *lacU169* ( $\Phi$ 80 *lacZAM15*)) was used as the recipient strain for cloning experiments and plasmid propagation. It was grown following standard procedures.

*S. cerevisiae* strain Y06460 (Euroscarf) (BY4741; Mat a; his3D1; leu2D0; met15D0; ura3D0; YMR318c::kanMX4) and derivatives were grown in yeast extract peptone dextrose media, Delft medium supplemented with sugar beet molasses and appropriate synthetic complete (SC) media.

Agrobacterium strain AGL1 was used for the transient and stable expression assays *in planta* and grown following standard procedures in LB medium with appropriate antibiotics.

**Transcriptomic analysis.** Total RNA from *V. planifolia* was prepared from ~100 mg of 6-month-old vanilla pod tissue using the RNeasy plant mini kit (Qiagen, <http://www.qiagen.com>) with on-column DNase I digestion. Quality of RNA was assessed using a bioanalyzer (Agilent). RNA (about 100  $\mu\text{g}$  total) was provided to Eurofins MWG Operon ([www.eurofinsdna.com](http://www.eurofinsdna.com)) for 454 Roche sequencing. A normalized assembly of sequences was obtained. Obtained sequencing reads were *de novo* assembled using CLC Genomics Workbench 5.0 with default settings. Raw sequencing reads were submitted to the Sequence Read Archive (SRA) database at National Center for Biotechnology Information (Bioproject accession no. SRP023166). Total RNA from *Glechoma hederacea* was isolated using the RNeasy plant kit (Qiagen) and provided to Macrogen ([www.macrogen.com](http://www.macrogen.com)) for Illumina HiSeq sequencing.

Raw sequencing reads were submitted to the Sequence Read Archive database at the National Center for Biotechnology Information (accession no. KJ775791).

**Proteomic analysis.** The inner part of the vanilla pod was ground to a fine powder in liquid nitrogen and extracted in 400 mM Tris/HCl (pH 8), 20 mM  $\text{MgCl}_2$ . Solubilized proteins were separated by SDS-PAGE on 10% Tris-glycine gels (Bio-Rad) and stained using Coomassie Brilliant Blue R-250 (Bio-Rad). Protein masses were estimated using a standard broad range Bio-Rad molecular mass marker. In-gel digestion of protein bands and MS were performed by the University of Victoria—Genome BC Proteomic center (method information found on [www.proteincentre.com](http://www.proteincentre.com))<sup>48</sup>. The peptide mass data and tandem mass data obtained were used to search known protein sequences from the *V. planifolia* transcriptome data.

**Isolation and subcloning of genes.** A complementary DNA (cDNA) library made from a 6-month-old *V. planifolia* pod was kindly provided by Evolve A/S Denmark. The cDNA library was inserted in a pYES2 vector (Invitrogen) ([http://tools.invitrogen.com/content/sfs/manuals/pyes2\\_man.pdf](http://tools.invitrogen.com/content/sfs/manuals/pyes2_man.pdf)). cDNA from *Glechoma hederacea* was made from material sourced in Basel, Switzerland. Total RNA was

isolated using the RNeasy plant kit (Qiagen) and cDNA was made using the Mint2 cDNA synthesis kit (Evrogen) ([www.evrogen.com](http://www.evrogen.com)).

Candidate genes identified from the transcriptome data were amplified from the cDNA library by PCR with gene-specific primers (Supplementary Tables 1 and 3) to obtain full-length sequences. The PCR products were subsequently cloned in blunt-II-topo vector (Invitrogen) in *E. coli*. Plasmids were purified using the miniprep kit (Qiagen) and gene sequences were confirmed by sequencing.

**Construction of expression cassettes.** Yeast expression plasmids were constructed by transferring gene inserts by restriction digestion cloning with the suitable restriction enzymes and subsequently ligated into the yeast expression vector p426-GPD, containing the constitutive *GPD* promoter and the p416-TEF vector containing the constitutive *TEF* promoter<sup>49</sup> (Supplementary Table 1). Restriction enzymes and T4 ligase were from New England Biolabs and reactions were carried out according to manufacturers' instructions. Vectors for chromosomal integration of genes were constructed using the uracil-specific excision reagent (USER) cloning method and a vector system adapted from Mikkelsen *et al.*<sup>50</sup> A plasmid (pVAN714) containing UP- and DOWN-targeting fragments for replacing the yeast endogenous exoglucosidase 1 gene (*EXG1*) was constructed similarly as described in Mikkelsen *et al.*<sup>50</sup> The genes and a fused *TEF1/PGK1* promoter DNA fragment were USER cloned into pVAN714.

Plant expression vectors for tobacco transient expression were constructed using Gateway cloning technology (Invitrogen). The cDNAs of interest were PCR amplified with gene-specific primers with *attB* overhangs; 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTAAAAATGTCTATGGCAGCTAAGCTCCTCTTC-3' and 5'-GGGGACCCAGCTTTCTTGATCAAAAGTGGTCACAGCCACAATGGATAAGATG-3' (All primers are listed in Supplementary Table 3) and cloned in the pDONR207 gateway vector (Invitrogen) and subsequently transferred to the destination vector pJAM1502 (ref. 51) by homologous recombination.

Plant expression vectors for tobacco transient expression to test the catalytic activity of *VpVAN* and different modifications thereof were constructed using the Gateway cloning technology (Invitrogen). The cDNAs of interest were PCR amplified with *attB* overhangs; *VpA61van* 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAAAAATGTCTTCGATGAGGTACGGGAAGAGCTACGGATCGGAG-3' and 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTACACAGCCACAATCAACAATGGGATAAG-3', *VpA137van* 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAAAAATGTCTTCGATGAGGTACGGGTCTGTAACGAGGGA-3' and 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTACACAGCCACAATGGGATAAG-3', and cloned in the pDONR207 gateway vector (Invitrogen) and subsequently transferred to the destination vector pEAQ-HT-DEST3 (ref. 52) by homologous recombination. *vpnbAspA137van* (sequence: Supplementary Fig. 2e) was synthetically synthesized by MWG-Biotech ([www.mwg-biotech.com](http://www.mwg-biotech.com)). Empty vector control was the destination vector pEAQ-HT-DEST3 (ref. 52) having a nonsense gene with a few amino acids having a start and a stop codon.

The binary vector pUCEUBI:SP-USER:NOS was used for stable transformation of *Barley*. Engineering of the construct was done as described previously<sup>53</sup>. In short, *VpVAN* was amplified by PCR using the specific primers: *VpHvVAN* 5'-GGTCTTAAUATGGCAGCTAAGCTCCTC-3' and 5'-GGCATTAATCAAA CAGCCACAATGGGGTATG-3' and *VpHvAspvan* 5'-GGTCTTAAUATGTTCCTGCTGTTTCTAGTGTCCG-3' and 5'-GGCATTAATCAAAACAGCCACAATGGGGTATG-3'. The PCR-reactions were carried out using PfuTurbo CX Hotstart DNA polymerase (Stratagene) according to the manufacturer's instructions. The resulting PCR product was inserted into the binary vector by using USER-TM cloning as described previously<sup>53</sup>.

**In vitro transcription/translation.** The TNT Quick Coupled Transcription/Translation kit for PCR-generated DNA (Promega) was used to produce proteins of interest directly from PCR products. L-[<sup>35</sup>S]-Methionine was included to permit monitoring of the radiolabelled proteins formed following separation by SDS-PAGE and visualized by incubating dried gels for 48 h on phosphorimager screens, which were scanned with a STORM 860 molecular imager (Molecular Dynamics).

**In vitro protein assay.** Proteins produced in coupled *in vitro* transcription/translation assays were analysed for their enzyme catalytic abilities by incubation of aliquots (10  $\mu\text{l}$ ) with 0.5–5 mM of the following substrates: ferulic acid (Sigma), *p*-coumaric acid (Sigma), caffeic acid (Sigma), ferulic acid glucoside, *p*-coumaric acid glucoside, caffeic acid glucoside, caffeoyl-Coenzyme A (MicroCombiChem e.K.), *p*-coumaryl-Coenzyme A (MicroCombiChem e.K.), feruloyl-Coenzyme A (MicroCombiChem e.K.) or sinapyl-Coenzyme A (MicroCombiChem e.K.) in 400 mM Tris/HCl (pH 8), 20 mM  $\text{MgCl}_2$  and 2.5 mM dithiothreitol (total volume: 50  $\mu\text{l}$ ). Enzyme assays were carried out in the presence and absence of 2.5 mM dithiothreitol, 0.1 mM ATP and 0.1 mM  $\text{NAD}^+$ . Aliquots (10  $\mu\text{l}$ ) were withdrawn at specific time points and enzyme activity stopped by MeOH addition (25  $\mu\text{l}$ , 25% (v/v)) and heating (45 °C, 15 min). Samples were cooled on ice (30 min), centrifuged (7,400 g, 10 min) in microtitre filter plates (Merck Millipore) and the filtrate was finally analysed by LC-MS.

**Yeast transformation.** Yeast expression plasmids with candidate genes were transformed into *S. cerevisiae* using the LiAc/SS carrier DNA/polyethylene glycol



method<sup>54</sup>. All transformants were grown on SC medium lacking Uracil (URA) to select positive transformants (yeast strains constructed—Supplementary Table 2).

**Enzyme assays using yeast.** Transformed yeast cells were cultured in 50 ml of liquid SC-URA for 24 h. Putative substrates (2.5 mM) were administered to the yeast culture (3 ml) and incubated (28 °C, 72 h, 200 r.p.m.) using a sterile 24-well microtitre plate (Biopioneer Inc). Incubation was stopped by addition of MeOH (350 µl, 35% (v/v)) and the samples prepared for LC–MS as described above.

**Transient expression of VpVAN in tobacco.** Overnight cultures of an *Agrobacterium tumefaciens* strain AGL1 containing either the recombinant pJAM1502 or pEAQ-HT-DEST3 vectors harbouring the VpVAN cDNA and an *A. tumefaciens* strain AGL1 carrying the recombinant pJAM1502 vector harbouring the gene-silencing inhibitor protein 19 (p19)<sup>34</sup> were harvested by centrifugation and resuspended (OD<sub>600</sub> = 2.0) in 10 mM MES pH 5.5, 10 mM MgCl<sub>2</sub> and 100 µM acetosyringone. After incubation (4 h, room temperature), the two *A. tumefaciens* strains were used to co-infiltrate leaves of 3-week-old *N. benthamiana* plants grown at 24 °C (day) and 17 °C (night). After 4 or 5 days, leaf discs (1 cm diameter) were stamped out from the infiltrated leaves and metabolites extracted in 60% (v/v) MeOH for LC–MS analysis.

**Stable expression of VpHvVAN in barley.** The vectors harbouring VpHvVAN and VpHvAspvan cDNA (as described previously) were transformed into *A. tumefaciens* strain AGL0 using the freeze/thaw method and selected on medium with 50 mg l<sup>-1</sup> spectinomycin and 25 mg l<sup>-1</sup> rifampicin. Immature embryos isolated from barley plants 12–14 days after pollination were used for *Agrobacterium*-mediated transformation following the procedure described elsewhere<sup>55</sup>. Regenerated transgenic plants were transferred to the greenhouse. Leaf samples were collected 6–8 weeks after transfer to the greenhouse.

**Biosynthetic assays with green vanilla pod discs.** Fresh vanilla pods were harvested 3 and 6 months after pollination. The pods were cut into small discs using a scalpel and further dissected to separate the inner and outer part of the pod. Radiolabelled precursors (0.5 µCi) were administered to samples representing the inner and outer part of the pod (approx. identical fresh weight) and embedded (30 °C) in 400 mM Tris/HCl pH 8, 20 mM MgCl<sub>2</sub> for 24 h.

**In tube in situ PCR using tissues of vanilla pod discs.** Fresh vanilla pods from *V. planifolia* were cut into small pieces (2–4 mm<sup>3</sup>) and immediately fixed (4 h, 4 °C) in freshly prepared aqueous FAA (2% (v/v) formaldehyde, 5% (v/v) acetic acid, 63% (v/v) ethanol in phosphate-buffered saline). The transcript level of VpVAN in different cell types was visualized based on specific primers<sup>56</sup>; *in situ* VpVAN 5'-AAGCCTTGAATACGTTAAGTACAATGGA-3' and *in situ* VpVAN reverse 5'-GTGTCACTGCTGTATACACCTTCTT-3'.

**Analytical chemistry.** The <sup>14</sup>C-labelled products formed in biosynthetic experiments with fresh vanilla pods as well as in *in vitro* protein assays were applied to Silica Gel 60 F254 TLC plates (Merck, <http://www.merck-chemicals.com>). The plates were developed in ethyl acetate: acetone: dichloromethane: methanol: water (40:30:12:10:8, v/v/v/v/v), dried, exposed (48 h) on phosphorimager screens (Molecular Dynamics, <http://www.molecular-dynamics.com>) and the radiolabelled compounds formed were visualized using a Storm 860 Molecular Imager (Molecular Dynamics). Identification of the radiolabelled compounds formed was guided by co-application of authentic standards. Unambiguous structural verification of the products formed was obtained using LC–MS including accurate mass determination and comparison of retention times and fragmentation patterns with those of authentic reference compounds<sup>57</sup>.

**Chemical synthesis.** *p*-Hydroxybenzaldehyde glucoside and vanillyl alcohol glucoside: the corresponding aglycons were glucosylated using 2,3,4,6-tetra-*O*-acetyl- $\alpha$ -D-glucopyranosyl bromide. The reaction was performed in aqueous organic basic media using homogeneous reaction conditions and aqueous NaOH with acetone as the organic co-solvent following the method reported by Mauthner<sup>58</sup>. The target glucosides were obtained by Zemplén deacylation of the aryl *O*-protected glucosides.

Vanillic acid glucoside: vanillin *O*-protected glucoside was synthesized as mentioned for *p*-hydroxybenzaldehyde glucoside and oxidized to the corresponding carboxylic acid using potassium permanganate (KMnO<sub>4</sub>). Deacetylation was accomplished as reported above.

4- $\beta$ -D-glucopyranosylcoumaric acid: *p*-coumaric acid glucoside and ferulic acid glucoside were chemically synthesized according to Galland *et al.*<sup>59</sup> and references therein.

Ferulic acid glucoside-<sup>13</sup>C/<sup>14</sup>C<sub>6</sub> analogue: to synthesize the desired ferulic acid-<sup>13</sup>C/<sup>14</sup>C<sub>6</sub>- $\beta$ -D-glucoside, the key step is the glycosylation of methyl ferulate by 1-fluoro-2,3,4,6-tetra-*O*-acetyl- $\alpha$ -D-glucopyranose-<sup>13</sup>C/<sup>14</sup>C<sub>6</sub> with the BF<sub>3</sub>·Et<sub>2</sub>O complex as the activator<sup>60</sup>. In this work, D-glucose-13C<sub>6</sub>/14C<sub>6</sub> in the ratio

of 98:2 was converted into the corresponding glycosyl fluoride as reported previously<sup>61</sup>.

The purity and structural conformation of the synthesized compounds were verified by NMR spectroscopy. In all cases, the <sup>1</sup>H- and <sup>13</sup>C-NMR chemical shifts for the chemically synthesized target molecules were consistent with previously reported data: *p*-coumaric acid glucoside (4- $\beta$ -D-glucopyranosylcoumaric acid) and ferulic acid glucoside (4- $\beta$ -D-glucopyranosylferulic acid) as reported by Galland *et al.*<sup>59</sup> and references therein, *p*-hydroxybenzaldehyde glucoside (4- $\beta$ -D-glucopyranosylbenzaldehyde)<sup>62</sup>, vanillyl alcohol glucoside (4- $\beta$ -D-glucopyranosylvanillyl alcohol)<sup>63</sup> and vanillic acid glucoside (4- $\beta$ -D-glucopyranosylvanillic acid)<sup>64</sup>.

## References

- Sharp, M. D., Kocaoglu-Vurma, N. A., Langford, V., Rodriguez-Saona, L. E. & Harper, W. J. Rapid discrimination and characterization of vanilla bean extracts by attenuated total reflection infrared spectroscopy and selected ion flow tube mass spectrometry. *J. Food Sci.* **77**, C284–C292 (2012).
- Radulovic, N., Dordevic, N., Markovic, M. & Palic, R. Volatile constituents of *Glechoma hirsuta* Waldst. & Kit. And *G. hederacea* L. (Lamiaceae). *Bull. Chem. Soc. Ethiop.* **24**, 67–76 (2010).
- Sinha, A. K., Sharma, U. K. & Sharma, N. A comprehensive review on vanilla flavor: Extraction, isolation and quantification of vanillin and others constituents. *Int. J. Food Sci. Nutr.* **59**, 299–326 (2008).
- Walton, N. J., Mayer, M. J. & Narbad, A. Molecules of interest—vanillin. *Phytochemistry* **63**, 505–515 (2003).
- Lesage-Meessen, L. *et al.* Fungal transformation of ferulic acid from sugar beet pulp to natural vanillin. *J. Sci. Food Agr.* **79**, 487–490 (1999).
- Hansen, E. H. *et al.* *De novo* biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). *Appl. Environ. Microb.* **75**, 2765–2774 (2009).
- Bennett, D. J. & Kirby, G. W. Constitution and biosynthesis of capsaicin. *J. Chem. Soc.* **C** 442–446 (1968).
- Funk, C. & Brodelius, P. Phenylpropanoid metabolism in suspension cultures of *Vanilla planifolia*. I. Influence of growth regulators and an elicitor on phenylpropanoid metabolism in suspension cultures of *Vanilla planifolia*. *Phytochemistry* **29**, 845–848 (1990).
- Funk, C. & Brodelius, P. E. Phenylpropanoid metabolism in suspension cultures of *Vanilla planifolia* Andr. II. Effects of precursor feeding and metabolic inhibitors. *Plant Physiol.* **94**, 95–101 (1990).
- Funk, C. & Brodelius, P. E. Phenylpropanoid metabolism in suspension cultures of *Vanilla planifolia* Andr. III. Conversion of 4-methoxycinnamic acids into 4-hydroxybenzoic acids. *Plant Physiol.* **94**, 102–108 (1990).
- Hahlbrock, K. & Scheel, D. Physiology and molecular biology of phenylpropanoid metabolism. *Annu. Rev. Plant Phys. Plant Mol. Biol.* **40**, 347–369 (1989).
- Daphna Havkin-Frenkel, F. B. *Handbook of Vanilla Science and Technology* Ch. 17–18 (Wiley-Blackwell, 2010).
- Anwar, M. H. Paper chromatography of monohydroxyphenols in vanilla extract. *Anal. Chem.* **35**, 1974–1976 (1963).
- Boatright, J. *et al.* Understanding *in vivo* benzenoid metabolism in petunia petal tissue. *Plant Physiol.* **135**, 1993–2011 (2004).
- Schnitzler, J. P., Madlung, J., Rose, A. & Seitz, H. U. Biosynthesis of *p*-hydroxybenzoic acid in elicitor-treated carrot cell-cultures. *Planta* **188**, 594–600 (1992).
- Zenk, M. H. Biosynthese von vanillin in *Vanilla planifolia* Andr. *Z. Pflanzenphysiologie* **53**, 404–414 (1965).
- Van Moerkercke, A., Schavvinhold, I., Pichersky, E., Haring, M. A. & Schuurink, R. C. A plant thiolase involved in benzoic acid biosynthesis and volatile benzenoid production. *Plant J.* **60**, 292–302 (2009).
- Havkin-frenkel, D., Podstolski, A. & Dixon, R. A. Vanillin Biosynthetic Pathway Enzyme from *Vanilla planifolia*. US Patent US20030070188 A1 (US Patent Office: United States, 2003).
- Podstolski, A. *et al.* Unusual 4-hydroxybenzaldehyde synthase activity from tissue cultures of the vanilla orchid *Vanilla planifolia*. *Phytochemistry* **61**, 611–620 (2002).
- Fritz, R. R., Hodgins, D. S. & Abell, C. W. Phenylalanine ammonia-lyase. Induction and purification from yeast and clearance in mammals. *J. Biol. Chem.* **251**, 4646–4650 (1976).
- Ro, D. K., Mah, N., Ellis, B. E. & Douglas, C. J. Functional characterization and subcellular localization of poplar (*Populus trichocarpa*  $\times$  *Populus deltoides*) cinnamate 4-hydroxylase. *Plant Physiol.* **126**, 317–329 (2001).
- Schoch, G. *et al.* CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. *J. Biol. Chem.* **276**, 36566–36574 (2001).
- Lam, K. C., Ibrahim, R. K., Behdad, B. & Dayanandan, S. Structure, function, and evolution of plant *O*-methyltransferases. *Genome* **50**, 1001–1013 (2007).
- Odoux, E. & Brillouet, J. M. Anatomy, histochemistry and biochemistry of glucovanillin, oleoresin and mucilage accumulation sites in green mature



- vanilla pod (*Vanilla planifolia*; Orchidaceae): a comprehensive and critical reexamination. *Fruits* **64**, 221–241 (2009).
25. Fan, J. *et al.* Expression of a senescence-associated cysteine protease gene related to peel pitting of navel orange (*Citrus sinensis* L. Osbeck). *Plant Cell Tiss. Organ Cult.* **98**, 281–289 (2009).
  26. Jorgensen, K. *et al.* Biosynthesis of the cyanogenic glucosides linamarin and lotaustralin in cassava: isolation, biochemical characterization, and expression pattern of CYP71E7, the oxime-metabolizing cytochrome P450 enzyme. *Plant Physiol.* **155**, 282–292 (2011).
  27. Huang, Z. X., Dostal, L. & Rosazza, J. P. N. Microbial transformations of ferulic acid by *Saccharomyces cerevisiae* and *Pseudomonas fluorescens*. *Appl. Environ. Microb.* **59**, 2244–2250 (1993).
  28. Gu, W. *et al.* Structural basis of enzymatic activity for the ferulic acid decarboxylase (FADase) from *Enterobacter* sp Px6-4. *PLoS ONE* **6**, e16262 (2011).
  29. Mukai, N., Masaki, K., Fujii, T., Kawamukai, M. & Iefuji, H. PAD1 and FDC1 are essential for the decarboxylation of phenylacrylic acids in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* **109**, 564–569 (2010).
  30. Payet, B., Shum Cheong Sing, A. & Smadja, J. Comparison of the concentrations of phenolic constituents in cane sugar manufacturing products with their antioxidant activities. *J. Agric. Food Chem.* **54**, 7270–7276 (2006).
  31. Valli, V. *et al.* Sugar cane and sugar beet molasses, antioxidant-rich alternatives to refined sugar. *J. Agric. Food Chem.* **60**, 12508–12515 (2012).
  32. Schroeder, C., Sommer, J., Humpfer, E. & Stockigt, J. Inverse correlated H-1-C-13 in vivo NMR as a probe to follow the metabolism of unlabeled vanillin by plant cell cultures. *Tetrahedron* **53**, 927–934 (1997).
  33. Karen Wilson, D. M. *Monocots: Systematics and Evolution* (CSIRO PUBLISHING, 2000).
  34. Ibdah, M., Chen, Y. T., Wilkerson, C. G. & Pichersky, E. An aldehyde oxidase in developing seeds of *Arabidopsis* converts benzaldehyde to benzoic acid. *Plant Physiol.* **150**, 416–423 (2009).
  35. Gasson, M. J. *et al.* Metabolism of ferulic acid to vanillin. A bacterial gene of the enoyl-SCoA hydratase/isomerase superfamily encodes an enzyme for the hydration and cleavage of a hydroxycinnamic acid SCoA thioester. *J. Biol. Chem.* **273**, 4163–4170 (1998).
  36. Mitra, A. *et al.* 4-Hydroxycinnamoyl-CoA hydratase/lyase, an enzyme of phenylpropanoid cleavage from *Pseudomonas*, causes formation of C(6)-C(1) acid and alcohol glucose conjugates when expressed in hairy roots of *Datura stramonium* L. *Planta* **215**, 79–89 (2002).
  37. Leonard, P. M. *et al.* The 1.8 angstrom resolution structure of hydroxycinnamoyl-coenzyme A hydratase-lyase (HCHL) from *Pseudomonas fluorescens*, an enzyme that catalyses the transformation of feruloyl-coenzyme A to vanillin. *Acta Crystallogr. D* **62**, 1494–1501 (2006).
  38. Wong, B. J. & Gerlt, J. A. Evolution of function in the crotonase superfamily: (3S)-methylglutaconyl-CoA hydratase from *Pseudomonas putida*. *Biochemistry* **43**, 4646–4654 (2004).
  39. Achterholt, S., Priefert, H. & Steinbuchel, A. Identification of Amycolatopsis sp. strain HR167 genes, involved in the bioconversion of ferulic acid to vanillin. *Appl. Microbiol. Biotechnol.* **54**, 799–807 (2000).
  40. Cambra, I., Hernandez, D., Diaz, I. & Martinez, M. Structural basis for specificity of propeptide-enzyme interaction in barley C1A cysteine peptidases. *PLoS ONE* **7**, e37234 (2012).
  41. Turk, V. *et al.* Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim. Biophys. Acta* **1824**, 68–88 (2012).
  42. Karrer, K. M., Peiffer, S. L. & DiTomas, M. E. Two distinct gene subfamilies within the family of cysteine protease genes. *Proc. Natl Acad. Sci. USA* **90**, 3063–3067 (1993).
  43. Matarasso, N., Schuster, S. & Avni, A. A novel plant cysteine protease has a dual function as a regulator of 1-aminocyclopropane-1-carboxylic Acid synthase gene expression. *Plant Cell* **17**, 1205–1216 (2005).
  44. Visal, S., Taylor, M. A. & Michaud, D. The proregion of papaya proteinase IV inhibits Colorado potato beetle digestive cysteine proteinases. *FEBS Lett.* **434**, 401–405 (1998).
  45. Marra, B. M. *et al.* Protective effects of a cysteine proteinase propeptide expressed in transgenic soybean roots. *Peptides* **30**, 825–831 (2009).
  46. Menard, R. & Storer, A. C. Oxyanion hole interactions in serine and cysteine proteases. *Biol. Chem. Hoppe Seyler* **373**, 393–400 (1992).
  47. Crespo, J. Compositions intended to increase piglet appetite. European patent WO2009068622-A12009 (European Patent Office, 2009).
  48. Borchers, C., Peter, J. F., Hall, M. C., Kunkel, T. A. & Tomer, K. B. Identification of in-gel digested proteins by complementary peptide mass fingerprinting and tandem mass spectrometry data obtained on an electrospray ionization quadrupole time-of-flight mass spectrometer. *Anal. Chem.* **72**, 1163–1168 (2000).
  49. Mumberg, D., Muller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156**, 119–122 (1995).
  50. Mikkelsen, M. D. *et al.* Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.* **14**, 104–111 (2012).
  51. Luo, J. *et al.* Convergent evolution in the BAHF family of acyl transferases: identification and characterization of anthocyanin acyl transferases from *Arabidopsis thaliana*. *Plant J.* **50**, 678–695 (2007).
  52. Peyret, H. & Lomonosoff, G. P. The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. *Plant Mol. Biol.* **83**, 51–58 (2013).
  53. Hebelstrup, K. H. *et al.* UCE: a uracil excision (USER (TM))-based toolbox for transformation of cereals. *Plant Methods* **6**, 15 (2010).
  54. Gietz, R. D. & Schiestl, R. H. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 35–37 (2007).
  55. Holme, I. B. *et al.* Cisgenic barley with improved phytase activity. *Plant Biotechnol. J.* **10**, 237–247 (2012).
  56. Jorgensen, K. *et al.* Cassava plants with a depleted cyanogenic glucoside content in leaves and tubers. Distribution of cyanogenic glucosides, their site of synthesis and transport, and blockage of the biosynthesis by RNA interference technology. *Plant Physiol.* **139**, 363–374 (2005).
  57. Saito, S., Motawia, M. S., Olsen, C. E., Moller, B. L. & Bak, S. Biosynthesis of rhodiocyanosides in *Lotus japonicus*: Rhodiocyanoside A is synthesized from (Z)-2-methylbutanaloxime via 2-methyl-2-butenenitrile. *Phytochemistry* **77**, 260–267 (2012).
  58. Mauthner, F. Über neue synthetische glucoside. *J. Prakt. Chem.* **85**, 564–568 (1912).
  59. Galland, S., Mora, N., Abert-Vian, M., Rakotomanomana, N. & Dangles, O. Chemical synthesis of hydroxycinnamic acid glucosides and evaluation of their ability to stabilize natural colors via anthocyanin copigmentation. *J. Agric. Food Chem.* **55**, 7573–7579 (2007).
  60. Kunz, H. & Sager, W. Stereoselective glycosylation of alcohols and silyl ethers using glycosyl fluorides and boron-trifluoride etherate. *Helv. Chim. Acta* **68**, 283–287 (1985).
  61. Motawia, M. S., Olsen, C. E., Moller, B. L. & Marcussen, J. Chemical synthesis and nmr-spectra of a protected branched-tetrasaccharide thioglycoside, a useful intermediate for the synthesis of branched oligosaccharides. *Carbohydr. Res.* **252**, 69–84 (1994).
  62. Manguro, L. O. & Lemmen, P. Phenolics of *Moringa oleifera* leaves. *Nat. Prod. Res.* **21**, 56–68 (2007).
  63. Koike, Y. *et al.* Novel phenolic glycosides, adenophorases A-E, from *Adenophora* roots. *J. Nat. Med.* **64**, 245–251 (2010).
  64. Sakushima, A., Coskun, M. & Maoka, T. Hydroxybenzoic acids from *Boreava orientalis*. *Phytochemistry* **40**, 257–261 (1995).

## Acknowledgements

This work was supported by a grant from the Danish Council for Strategic Research (DSF), by a grant to the VILLUM research center ‘Plant Plasticity’ and by the Center for Synthetic Biology: bioSYNergy supported by the UCPH Excellence Programme for Interdisciplinary Research. Jørgen Hansen (Evolva A/S, Denmark) is acknowledged for comments and corrections to the manuscript. We thank BRC, VATEL, CIRAD, La Réunion for providing *Vanilla planifolia* materials. We thank Hadrien Peyret and George Lomonosoff for providing pEAQ-HT-DEST3 vectors for plant gateway cloning.

## Author contributions

N.J.G. performed the [<sup>14</sup>C]-radiolabelled precursor feeding experiments and TLC analysis, mRNA extractions for 454 transcriptome sequencing and handling 454 sequence data, proteomic studies and handling proteomic data, *in vitro* transcription/translation studies, all the molecular biology analysis, yeast strain constructions, tobacco expression studies, in tube *in situ* PCR and contributed to writing the manuscript. E.H.H. constructed all the yeast integration plasmids, contributed to planning, designing the project and writing the manuscript. R.K. contributed to performing the proteomic studies and writing the manuscript. C.E.O. carried out the LC-MS analysis. M.S.M. designed and performed the chemical synthesis of all glucoside substrates. K.J. sectioned the vanilla pod for in tube *in situ* PCR and performed the fluorescence microscopy. M.G. contributed to writing the manuscript and was in charge of courier shipment of fresh vanilla materials from La Réunion to Denmark. I.H. and K.H. contributed to the planning and performance of the barley transformation. B.L.M. planned and designed the project, provided biochemical expertise and scientific mentoring and contributed to writing the manuscript.

## Additional information

**Accession codes:** Raw sequencing reads of *V. planifolia* transcriptome is submitted to the Sequence Read Archive (SRA) database at National Center for Biotechnology Information (Bioproject accession no. SRP023166). The raw sequencing data reads have been deposited in the Sequence Read Archive under accession code KJ775791.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The biotech company Evolva A/S has developed a yeast-based production platform for synthesis of vanillin glucoside from glucose (ref. 6). Esben H. Hansen is a research employee at Evolva and Birger Lindberg Møller is on the Science Advisory Board of Evolva. The remaining authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Gallage, N. J. *et al.* Vanillin formation from ferulic acid in *Vanilla planifolia* is catalysed by a single enzyme. *Nat. Commun.* 5:4037 doi: 10.1038/ncomms5037 (2014).



This work is licensed under a Creative Commons Attribution 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>